

Reliability and Validity of Survey Data on Physical Health

JOSEPHINE WILLIAMS MELTZER, Ph.D., and JOSEPH R. HOCHSTIM, Ph.D.

HOW consistently do people answer questions about their health when a survey is repeated after a short interval? How closely does the information collected by survey agree with that obtained from clinical records? This paper evaluates survey data on physical health from these two points of view. The two methodological studies discussed are part of a research program conducted by the Human Population Laboratory of the California State Department of Public Health and supported by grants from the Public Health Service.

In 1965 health data were collected by self-administered questionnaires from a large probability sample of adults in Alameda County, Calif. Known as the Survey of Health and Ways of Living, this baseline survey is the first of a projected series in a longitudinal study of physical, mental, and social health in the county. The study's purpose is to assess the relation between health, understood in a broad generic sense, and a wide range of demographic, economic, familial, cultural, and environmental factors and personal habits, which for convenience are referred to collectively as "ways of living."

A primary hypothesis in this study is that in

Dr. Meltzer and Dr. Hochstim are with the Human Population Laboratory, California State Department of Public Health, Berkeley. This study was supported by research grant HS00368 from the Division of Health Care Services, Public Health Service.

a general population physical health is positively associated with both mental and social health. To test this hypothesis, a measure was devised to summarize each respondent's report of his physical health. This measure, the Physical Health Spectrum, divides the sample into four broad groups ranging from those who report some degree of disability to those who report no chronic condition. One intermediate group consists of people who are not in any way disabled but who suffer from some chronic illness or impairment, and the other intermediate group consists of those whose only chronic complaints are symptoms.

To determine the reliability of the Physical Health Spectrum and other items and indices, the Human Population Laboratory conducted a reliability survey of 1,530 residents of Alameda County in 1968. (None of these respondents had participated in the 1965 baseline survey; the two samples were drawn independently.) Validity, or the extent of agreement between the survey information and that obtained from clinical records, was investigated separately—by a record check on respondents in the baseline survey who had been receiving medical care through the Kaiser Foundation Health Plan of Northern California.

We know of no comparable study of the reliability of response to a self-administered health questionnaire in a general population. The results of a small study in which 64 respondents

were reinterviewed after a week or two suggested that illnesses are reported with satisfactory reliability but some symptoms are not (1). Some of the research conducted by the Bureau of the Census to measure the quality of interviewing is also of interest in this connection (2).

Several major studies compared survey estimates of the prevalence of chronic diseases with estimates derived from clinical records. Balamuth (3) of the Health Insurance Plan of Greater New York noted that 44 percent of a selected list of chronic conditions reported by physicians were also reported in a household survey, and that 40 percent of the chronic conditions reported by survey could be matched with physicians' reports. Madow (4) of the Stanford Research Institute compared forms completed by physicians in the Southern California Permanente Medical Group with interview data collected by the Census Bureau; 55 percent of the medically diagnosed chronic conditions were reported in the interviews.

In the Hunterdon study, Trussel and Elinson (5) compared household survey data with diagnostic examinations conducted a year or so after the interview. They reported that 29 percent of the conditions found on examination had been reported by interview, and that nearly half of the conditions reported by interview were confirmed by subsequent examination.

We know that although survey data generally do not agree closely with medical records, the agreement is better for serious conditions, particularly those with a recent impact on the patient. The amount of agreement varies appreciably from one diagnostic category to another and depends on the study design—the wording of questions, the definition of a “match,” and the kind of information in the medical records. It has relatively little relation to the demographic characteristics of the respondents.

The Baseline Survey

A 23-page, self-administered questionnaire was constructed for the 1965 baseline survey. Almost all the questions were multiple-choice. The Physical Health Spectrum questions covered 33 specific complaints—five types of functional disability, 14 chronic conditions, three impairments, and 11 symptoms associated with

chronic illness. The respondents were also asked to list any other health problems they had. The following wording was used for specific complaints:

DISABILITIES

Here is a list of activities that people sometimes have trouble with: trouble feeding themselves, trouble dressing themselves, trouble moving around. Do *you* have trouble doing any of these things? (If “Yes”: How long have you had this trouble? Less than 6 months or more than 6 months?)

Here are two more activities that people sometimes have trouble with: trouble climbing stairs, and trouble getting outdoors. Do *you* have trouble doing either of these things? (If “Yes”: etc.)

Are you now unable to work because of some illness or injury? (If “Yes”: etc.)

Have you had to change the kind of work you used to do, or had to cut down on the number of hours you used to work because of some illness or injury? (If “Yes”: etc.)

Have you had to cut down or stop any other activity you used to do because of some illness or injury? (For example, you have to “take it easy”, or cut out some sport, or you find you can't spend as many hours gardening as you used to. (If “Yes”: etc.)

CHRONIC CONDITIONS

Here is a list of medical conditions that usually last for some time. Have you had any of these conditions during the past 12 months? High blood pressure? Heart trouble? Stroke? Chronic bronchitis? Asthma? Arthritis or rheumatism? Epilepsy? Diabetes? Cancer? Tuberculosis? Stomach ulcer or duodenal ulcer? Chronic gallbladder trouble? Chronic liver trouble? Hernia or rupture?

IMPAIRMENTS

Here is a list of physical impairments. Do you have any of these? Missing hand, arm, foot, or leg? Trouble with seeing (even with glasses)? Trouble with hearing (even with hearing aid)?

SYMPTOMS

Here is a list of physical ailments. Have you had any of these during the past 12 months? Frequent cramps in the legs? Pain in the heart or tightness or heaviness in the chest? Trouble breathing or shortness of breath? Swollen ankles? Pains in the back or spine? Repeated pains in the stomach? Frequent headaches? Constant coughing or frequent heavy chest colds? Paralysis of any kind? Stiffness, swelling, or aching in any joint or muscle? Getting very tired in a short time?

A series of experimental studies was conducted to determine the best way to collect data for the baseline survey (6). As a result, the following strategy was adopted. A sample of households was visited by enumerators who collected a limited amount of demographic

information and left questionnaires to be returned by mail for each adult in the household. (An "adult" was anyone at least 20 years old or a younger person who had ever been married.) This initial effort to enlist cooperation was followed by three others—first by letter, then by telegram, and finally by personal contact. In the final wave, the remaining nonrespondents were assigned to interviewers who were instructed to pick up as many more completed questionnaires as they could, offering nonrespondents whatever help they might need, with the understanding that the information should be obtained by interview as a last resort.

For the baseline survey, a probability sample of 4,735 addresses was drawn in Alameda County (7). The enumeration was completed at 97 percent of the occupied households, and questionnaires were eventually obtained from 86 percent (6,928) of the adults who had been enumerated.

The Reliability Survey

Three years later, in 1968, the Human Population Laboratory conducted a special survey to determine the reliability of responses to the key questions in the baseline survey. Were respondents selecting responses haphazardly, or would they answer consistently if the same questions were repeated?

The basic design of the reliability study was to replicate the essential features of the baseline survey as closely as possible and then to persuade the respondents to answer the same questions again after a suitable interval. The choice of interval between the first and second responses represented a compromise between two principles. On one hand, the longer the interval the more risk that "real" changes in a respondent's health—or in his information about his health—would occur, causing the response reliability to be underestimated. On the other hand, if too short an interval were selected, respondents might remember how they had decided to answer the questions that they found difficult and might make a conscious effort to be consistent. Any such tendency might reduce the number of discrepancies with the result that reliability would be overestimated. After some discussion, 1 week was selected as the

interval to aim for. In practice, of course, the interval varied from one respondent to another. The median elapsed time was 7.2 days, with an interquartile range of 4.8 days.

A short version of the original baseline survey questionnaire was used in the reliability survey—8 pages rather than 23. It contained a total of 90 questions; 42 were on physical health and included all the questions cited. Households were enumerated, and self-administered questionnaires were left for each adult member, in essentially the same manner as in the baseline survey except that enumerators arranged to call back within a day or so to pick up the completed questionnaires. Before leaving the house, they mentioned that the study had two parts. On their second visit, the enumerators explained that the second part of the questionnaire would be mailed within a few days and made an appointment to return in about a week to pick it up. So that respondents would be forewarned of the repetition, the enumerators then mentioned casually that the second part included some of the same questions. This statement was repeated in the letter covering the second schedule. Actually the two forms were identical except for a few introductory questions on the first page and some minor changes in format. (Only one side of the paper was used the first time, but both sides the second time. The color of the paper was different.)

The sample for the 1968 reliability survey was drawn systematically from an area probability sample of addresses in Alameda County which had been listed in 1961. No effort was made in 1968 to sample dwellings that had been constructed in the intervening years at new addresses; so the sample did not represent the whole adult population of the county as of 1968. Of the 1,814 adults who were enumerated in 1968, 84 percent answered both "parts" of the questionnaire, yielding a sample of 1,530.

The 1968 respondents were slightly older than the 1965 (baseline) respondents, and perhaps as a consequence somewhat lower in the proportion of persons who had been to college. In general, we have found that better educated respondents are more reliable (8), and thus we may be underestimating the reliability of the original survey. Members of the reliability sample were also more likely to report a physical disability

than were members of the baseline sample, as would be expected from the age difference. We are not sure that this can be explained wholly in terms of age, however, for the proportion who reported some degree of disability was relatively high in the reliability sample in each of four age groups.

Reliability findings. Each version of the questionnaire included the same 33 dichotomous questions on specific physical problems. Altogether, 96 percent of the original responses to these questions were repeated on the second form (table 1), which is *prima facie* evidence of a very satisfactory degree of reliability. However, measurement of reliability in terms of the percentage of responses repeated does not allow for the agreement that could be expected purely by chance. More than 90 percent of the answers to each questionnaire were "No," even though the particular chronic complaints covered by the survey were relatively common ones; this means that about five out of six answers (84

percent) would be expected to agree by chance alone. (The expected value in any cell of a contingency table with the same number of rows and columns is the value obtained from the two marginal distributions if there is no association whatsoever between the two variables.) The observed agreement of 96 percent thus exceeded the expected agreement of 84 percent by 12 percentage points.

For convenience of interpretation, the excess of the observed agreement over the expected agreement is expressed as a percentage of its maximum value and referred to as the index of reliability. In formula form,

$$\text{Index of reliability} = \frac{(\text{maximum agreement} - \text{expected agreement})}{(\text{observed agreement} - \text{expected agreement})} \times 100,$$

where both the expected agreement and the maximum agreement are computed from the two marginal distributions. (If the marginal distributions of the two questionnaires are

Table 1. Index of reliability and percentage of second responses in agreement with first responses for 1968 sample, by type of complaint

Type of complaint	Total number responses	Index of reliability	Percent responses in agreement with first response		
			All	Yes	No
All types.....	50, 186	82	96	75	99
Disabilities.....	7, 610	80	96	75	98
Chronic conditions.....	21, 334	89	99	86	100
Impairments.....	4, 571	82	98	71	99
Symptoms.....	16, 671	79	93	72	97

NOTE: The method used to calculate index values is shown in table 2.

Table 2. Calculation of index value of 80 for disabilities in table 1, according to number of responses

First questionnaire	Second questionnaire								
	Observed			Expected			Maximum		
	Yes	No	All	Yes	No	All	Yes	No	All
Yes.....	597	198	795	76	719	795	730	65	795
No.....	133	6, 682	6, 815	654	6, 161	6, 815	0	6, 815	6, 815
Total.....	730	6, 880	7, 610	730	6, 880	7, 610	730	6, 880	7, 610

$$\text{Index of reliability} = \frac{(597 + 6, 682) - (76 + 6, 161)}{(730 + 6, 815) - (76 + 6, 161)} = \frac{7, 279 - 6, 237}{7, 545 - 6, 237} = .80$$

identical, the maximum agreement is the total number of respondents.) The index assumes the value 0 when the two sets of data agree no better (and no worse) than would be expected by chance and takes the value 100 when the observed agreement is equal to the maximum possible, given the actual marginal distributions (table 2).

With this definition, the set of questions on physical complaints has an index of reliability of 82, meaning that the first and second answers agree better than could be expected by chance.

One would expect the more frequent answer, "No," to be more "reliable" than the minority answer, "Yes," simply because the few pairs of inconsistent answers are a much smaller proportion of the large group of "No's" than of the small group of "Yes's." Actually, the difference was marked—99 percent of the original "No's" but only 75 percent of the original "Yes's" were repeated on the second form (table 1).

The chronic illnesses with an index of reliability of 89 were more reliably reported than were the disabilities, impairments, and symptoms with indices of 80, 82, and 79 respectively (table 1). It may well be that because people have learned the diagnostic labels for their chronic illnesses from their physicians, they tend to report them consistently, whether or not the condition troubles them at the time they complete the questionnaire. Also, their responses to the more subjective questions on other types of complaints tend to be contingent on their day-to-day sense of physical well-being.

We had hoped that the Physical Health Spectrum, which divides the sample into four broad groups based on the most serious type of physical problem reported, would be more reliable than the answers to questions about specific complaints. The index of reliability for the Physical Health Spectrum is actually 78, however, or slightly less than the index of 82 for the set of 33 dichotomous questions on specific complaints. Nevertheless, 1,208 respondents, or 80 percent, fell into the same category of the Physical Health Spectrum whichever version of the questionnaire was used, although only 379 (25 percent) could be expected to do so purely on the basis of chance (table 3).

The respondents made fewer complaints the second time—the total number of physical complaints was 10 percent lower on the second questionnaire. The poorly defined symptoms registered the largest decline in reported prevalence: "constant coughing or frequent heavy chest colds" dropped from 94 respondents to 70, or by 25 percent; "getting very tired in a short time" by 20 percent; "frequent headaches" by 17 percent; "trouble with seeing (even with glasses)" by 16 percent; and having to cut down or stop any customary activity other than work by 14 percent. The only chronic illness that showed a comparable drop was chronic bronchitis, which fell off by 18 percent. In general, the drop was greater for symptoms and impairments (13 percent) than for disabilities and chronic conditions (6 percent).

Why the second questionnaire elicited fewer complaints than the first is a matter of speculation. If respondents had been less interested the second time, and had answered less carefully, one would expect more omissions on the second form and, in particular, a marked decline in the number of "other complaints," that is, those which had to be written in because they were not presented in checklists. The evidence on the tendency to skip questions is inconclusive because the interviewers were instructed to call omissions to the respondent's attention when they could, that is, if the respondent was at home when the questionnaire was picked up. There is no indication, however, that the second questionnaire was answered less carefully.

Table 3. Physical Health Spectrum for the reliability sample, first questionnaire by second questionnaire according to number of respondents

First questionnaire	All respondents	Second questionnaire			
		A	B	C	D
All respondents...	1, 508	316	387	356	449
A Disabilities.....	354	271	50	24	9
B Chronic conditions or impairments but no disability.....	381	34	300	26	21
C Symptoms only....	378	6	30	280	62
D No problem.....	395	5	7	26	357

The number of write-in complaints actually dropped less (3½ percent) than the number of checklisted complaints; so again there is no evidence that respondents failed to give as much attention to the second form. It is possible that some respondents lose their feeling of anonymity when a study includes a followup, and for that reason become less candid. Another possibility is that the original opportunity to report problems relieved some tension and made people feel a little better. The number of neurotic tendencies revealed in answers to psychological questions and the number of pessimistic answers to attitude questions also fell off—as though respondents were either less willing to report problems the second time or actually felt a little better.

Medical Record Check

Although the title of this paper refers to “validity,” the goal of the medical record check was actually more modest—to find out how well the health data collected in the baseline survey agreed with information in a specific set of clinical records for the same respondents and, in particular, how well the survey rating on the Physical Health Spectrum agreed with corresponding ratings based on medical records.

The two sets of data are compared without assuming that one is necessarily more valid than the other. Chronic illnesses are a partial exception. Since the sample was limited to patients who had had some kind of general physical examination at Kaiser, and since physicians feel an obligation to record any evidence of chronic illness that they find, the medical data on chronic illness are presumed to be more valid in general than the survey reports. However, the survey deals not only with chronic illnesses, but also with disabilities, impairments, and recurring symptoms. It would be hard, even in principle, to validate the symptoms reported in a survey by reference to medical records, for the patient is the ultimate authority on his own aches and pains, barring outright malingering—and there is little incentive to mangle in the survey situation. If the medical records show the same symptoms that are reported in the survey, the agreement is an indication of stability over time rather than validity. Much the same

can be said of disabilities and impairments, as these terms are used in the survey.

Design of the medical record check. The study design was dictated to a considerable extent by the availability of clinical records. Of the 6,928 respondents in the baseline survey, more than 1,000 belonged to the Kaiser Foundation Health Plan of Northern California. This is a prepayment plan which provides comprehensive medical care to a substantial number of persons in the San Francisco Bay Area through the Permanente Medical Group and various Kaiser Foundation hospitals.

Two physicians on the staff of the Human Population Laboratory abstracted the charts of Kaiser plan members who had participated in the 1965 baseline survey. (The record check was made possible by Dr. Morris Collen, director of medical methods of research, Permanente Medical Group; Dr. A. L. Baritell and Dr. Joseph Sender, former and current physicians in chief at the Kaiser Oakland Hospital; and Dr. Bernard Rhodes, physician in chief at Kaiser Hayward.) No special effort was made in connection with the Human Population Laboratory record check to standardize the basic medical data. The study is based on the records maintained by clinicians for clinical purposes. Problems of using medical records to confirm survey reports have been summarized by Madow (4). As he pointed out, physicians differ among themselves in the thoroughness of their examination, the threshold at which they diagnose a condition, the terminology they use, and what information they consider essential to record.

As expected, the abstractors found considerable differences among examining physicians in the amount of detail that they recorded. But nearly half the sample of records included reports of a “multiphasic” examination, which is an extensive, highly standardized screening program, pioneered by the Kaiser Foundation. It requires at least two visits. At the first visit, which takes about 3 hours, the patient fills out several questionnaires and takes a series of laboratory tests including an electrocardiogram, phonocardiogram, blood pressure, respirometry, chest X-ray, mammography (for women), various tests of hearing and the eyes (visual acuity, ocular tension, pupillary escape, retinal photograph), pain response, Achilles reflex, and blood

and urine chemistry including glucose tolerance. On the second visit, a physician who has a computer printout of the test results before him, gives a physical examination. In addition, women are advised to have a pelvic examination, and anyone over 40 years old is advised to report for a rectal examination.

The abstractors summarized the medical records without knowing how the patient had answered the survey questionnaire. On the basis of the written record, without consulting the attending physician, they attempted to answer 40 of the same questions that respondents had been asked, including all the questions on specific physical complaints. Unlike the respondents, who were offered a choice of "Yes" or "No" for each complaint, the abstractors had a third option, "probably yes," to cover borderline situations. In this analysis, "probably yes" is treated as "Yes." (The two physicians who did the abstracting conferred frequently with each other at the beginning of the process and continued to consult each other on difficult cases thereafter. Although the extent of their agreement was never formally assessed, they felt that they shared a common understanding of the basis for rating patients.)

Sample for the record check. In the Survey of Health and Ways of Living, 1,084 respond-

Table 4. Sex, age, and rating on the Physical Health Spectrum for baseline and medical record check samples, in percentages

Item	Baseline sample (N=6,928)	Medical record check sample (N=739) ¹
Males.....	46	44
Age group (years):		
Under 30.....	24	19
30-44.....	32	35
45-64.....	31	35
65 and over.....	13	11
Physical Health Spectrum:		
Disabilities ²	15	12
Chronic conditions or im-		
pairments but no dis-		
ability.....	28	33
Symptoms only.....	28	30
No problem.....	29	25

¹ Data from the survey.

² Disabilities of less than 6 months' duration are excluded.

ents (16 percent of the sample) reported membership in the Kaiser plan, and a search was made for the clinical records of this group. This investigation yielded some useful information on 902 patients, but their records were not necessarily complete. A recent survey of the Kaiser plan in the San Francisco Bay Area (9) showed why this was to be expected. Adult members used physicians who were not associated with the Permanente Medical Group for one-third of their home and office visits; some persons spent part of the year outside the area served by the plan. Employed men sometimes preferred to consult an industrial physician if it meant that they lost less time from work. New members sometimes continued for a while to use the physicians who had treated them previously. And dependents who had additional health insurance through their own employment sometimes used the Kaiser plan selectively, either for routine care of minor conditions or, alternatively, for major problems that would otherwise be costly.

To insure that the survey data were compared with reasonably complete records, the present analysis is based on the 739 records which included a general physical examination, either the traditional kind or a multiphasic screening examination, or both. Ideally, each patient would have been examined close to the time when he completed the survey questionnaire, but in practice, of course, the interval between the examination and the survey varied, and the abstractors had to interpret the survey year loosely, using their judgment about extrapolating information from earlier or later periods.

The record check sample is compared with the whole baseline sample in table 4. In both, women slightly outnumbered men. The record check sample was somewhat more concentrated in the middle-age range, from 30 to 64 years, than was the population of the county. Both samples' reports of their physical health yielded about the same distribution of ratings on the Physical Health Spectrum, although here again the Kaiser subsample was somewhat more concentrated in the middle part of the range.

Medical record check findings. Agreement between the survey data and the medical data is measured by an index parallel to the index of reliability—the excess of the observed

agreement over the expected agreement, expressed as a percentage of the maximum value of the excess, given the two marginal distributions.

Of the patients' responses to the 33 questions on specific physical complaints, 92 percent appeared in the medical records (table 5). As in the reliability study, negative answers were more likely to appear than were positive answers: 97 percent of the "No's" but only a third of the "Yes's" appeared. Both marginal distributions were heavily weighted with negative answers, with the result that 88 percent of the responses could be expected to agree by chance alone. The index of agreement was 37.

Survey responses to the 14 questions on chronic illness agreed better with the medical records than did responses on the other types of complaint. The index of agreement for chronic conditions was 52, and more than half (54 percent) of the chronic conditions reported in the survey were also noted in the records. (For comparison with other studies, the base of the percentage may be changed—of the chronic conditions noted by the medical abstractors, 48 percent were also reported by the patients). We have seen that patients report chronic illnesses more reliably than they report the other, more subjective complaints, and this may be one reason that chronic illnesses are more likely to appear in the medical records. But it is not the only reason—there are two kinds of internal evidence that physicians record their diagnoses of chronic illness more systematically than they record the other types of complaint under consideration.

Table 6. Number of complaints reported by the survey and by the medical abstractors and percentage of borderline ratings by abstractors, by type of complaint

Type of complaint	Number reported by survey	Reported by abstractors	
		Total number	Percent borderline
All types.....	1, 716	1, 354	31
Disabilities ¹	172	59	73
Chronic conditions..	401	451	23
Impairments.....	61	51	63
Symptoms.....	1, 082	793	30

¹ Disabilities of less than 6 months' duration are excluded.

The first kind of internal evidence concerns the reliability of the abstracting process when it is done from the written record without consulting the attending physicians. No formal test was made of the reliability of the abstractors' interpretation of the records, but their use of the borderline ("probably yes") category gives a clue. They were more confident of their interpretation of chronic conditions than of other types of complaints, in the sense that only 23 percent of their ratings of chronic illnesses were borderline in contrast to 73 percent of the disabilities and 63 percent of the impairments (table 6, third column).

The second kind of internal evidence concerns the number of complaints reported for two subgroups of patients. In general, respondents made more complaints by survey than their

Table 5. Index of agreement between survey data and medical record data for medical record check sample, by type of complaint

Type of complaint	Total number of responses	Index of agreement	Percent of record items in agreement with survey response		
			All	Yes	No
All types.....	24, 301	37	92	33	97
Disabilities ¹	3, 693	45	95	16	99
Chronic conditions.....	10, 322	52	96	54	98
Impairments.....	2, 207	31	96	28	98
Symptoms.....	8, 079	28	84	28	93

¹ Disabilities of less than 6 months' duration are excluded.

Table 7. Chronic conditions and symptoms in medical record sample, by source of data and type of examination given

Variable and source of data	Multi- phasic examina- tion	Tradi- tional examina- tion only
Number of patients.....	356	383
Number of chronic conditions per patient:		
Survey.....	0. 43	0. 65
Medical abstractors.....	. 56	. 66
Number of symptoms per patient:		
Survey.....	1. 32	1. 59
Medical abstractors.....	1. 39	. 78
Ratio of symptoms to chronic conditions:		
Survey.....	3. 1	2. 5
Medical abstractors.....	2. 5	1. 2

records showed (table 6)—2.32 per person as against 1.83. (This statement refers to complaints in the various checklists. Abstractors were more likely than the respondents to write in some "other" complaint. Since only one such "other" complaint per patient was coded, the total number cannot be reported.) In particular, disabilities were reported more frequently by patients than by abstractors. The only type of complaint for which this was not true was chronic conditions, which were reported more frequently in the medical record than in the survey. (This difference in the number of chronic conditions per patient can be attributed to the patients who had received a multiphasic examination, as shown in table 7.)

The extent to which symptoms were recorded in the clinical charts depended on the type of physical examination given the particular patient. The multiphasic examination included two extensive self-administered questionnaires, the "past medical history" covering conditions that had bothered the patient more than 1 year earlier, and the "interval history" covering the preceding 6 or 12 months. Both histories included some diagnostic questions as well as questions on symptoms, but the patients' reports of chronic illness were subject to confirmation by the examining physician.

Table 7 shows the number of chronic conditions and the number of symptoms reported in

the survey and in the medical abstracts for two groups of patients—those who had had a multiphasic examination and those who had had only a traditional examination. The two groups are not directly comparable, for the multiphasic group had fewer chronic illnesses and fewer chronic symptoms, but the ratios of symptoms to chronic conditions are comparable. For the patients whose records included a multiphasic examination, the ratio based on the medical abstracts (2.5) is nearly as high as indicated by the survey (3.1). For the balance of the sample, however, the abstracts showed only 1.2 symptoms per chronic disease whereas the survey showed 2.5. Thus the medical data on symptoms are comparable to the survey data only if the record includes a multiphasic examination.

Physical Health Spectrum. The survey ratings on the Physical Health Spectrum are cross-classified by the medical ratings in table 8. The unit in this table is a person rather than a condition.

The medical records showed as many patients (344) as did the survey (333) to be suffering from a serious condition, that is, from a disability of some kind or a chronic illness or an impairment, even though more such conditions were reported by survey (634) than by the abstractors (561), as shown in table 6. One reason is that respondents rated themselves partially disabled more frequently than the abstractors so rated them (table 6), but most respondents who considered themselves disabled also re-

Table 8. Physical Health Spectrum: survey ratings by abstractor's ratings, according to number of respondents, for the medical record check sample

Survey ratings	All respond- ents	Abstractors' ratings			
		A	B	C	D
All respondents.....	739	29	315	189	206
A Disabilities ¹	92	19	55	13	5
B Chronic conditions or or impairments but no disability.....	241	10	166	34	31
C Symptoms only.....	222	----	54	87	81
D No problem.....	184	----	40	55	89

¹ Disabilities of less than 6 months' duration are excluded.

ported a chronic illness and had been diagnosed as having some chronic illness.

Despite the inherent differences between the survey data and the information in medical charts, the survey ratings on the Physical Health Spectrum agreed with the abstractors' ratings for 361 respondents (49 percent). By chance alone, 29 percent of the ratings would be expected to agree, and the index of agreement is 34, or a third of the theoretical maximum value.

In the reliability study, survey ratings on the Physical Health Spectrum were less reliable for the oldest respondents than for other age groups, which is not surprising when one considers that they have more ailments, and thus more borderline decisions to make in responding to a questionnaire. Nevertheless, the survey ratings of the oldest respondents in the medical record check sample agreed better with ratings based on the medical data than those of younger respondents. The indices of agreement for the four age groups from youngest (under 30) to oldest (65 or older) were 28, 25, 31, and 40.

One reason for the better agreement in the oldest respondents' ratings is that their records are relatively complete. This could be inferred from the fact that, in general, the older respondents had belonged to the Kaiser plan longer and had used it more, or simply from the fact that the more severe health problems of the aged require a thorough workup. We do not need to rely on such inferences, however, for

internal evidence provides confirmation. The direction of the discrepancy between survey and medical ratings on the Physical Health Spectrum is shown in table 9, for record check sample members in four age groups and for 163 survey respondents who were excluded from the medical record check sample because no record was found of a general physical examination. Among this last group whose medical records are known to be incomplete, serious problems appeared more frequently in their survey reports than in their medical reports. This same pattern—of more serious problems found in the survey than in the medical records—is also characteristic of the younger members of the record check sample. Among patients 45 and over, however, one source was as likely as the other to produce more serious ratings, as if this group's medical records were relatively complete and detailed.

Conclusions

Substantive results have been reported on the reliability of two kinds of survey data on physical health—a set of 33 “Yes” or “No” questions on chronic physical complaints and ratings on the Physical Health Spectrum, which classifies each respondent into one of four groups depending on the most serious type of complaint reported.

In both instances the results are gratifying. The index of reliability for the whole set of dichotomous questions on complaints was 82 on

Table 9. Physical Health Spectrum: direction of discrepancy between survey and medical ratings for medical record sample, by age groups, and for other respondents

Sample and age group (years)	Number respondents	Physical Health Spectrum		
		Percent survey rating more serious	Percent ratings agree	Percent medical rating more serious
Total record check sample.....	739	29	49	22
Under 30.....	138	38	47	15
30-44.....	262	31	46	23
45-64.....	257	26	49	25
65 and over.....	82	23	57	20
Other respondents ¹	163	42	46	12

¹ Excluded from the medical record check sample because their records did not include a general physical examination. As a group, they were somewhat younger than the medical record check sample—66 percent under 45 years in contrast to 54 percent.

a scale of 100, with zero corresponding to the agreement to be expected by chance. The questions on chronic illnesses were even more reliable (index of 89). The index of reliability of ratings on the fourfold Physical Health Spectrum was 78.

In our record check of survey data on the prevalence of chronic disease, about half (54 percent) of the chronic conditions reported by survey appeared in the clinical records. The index of agreement is 52 on a scale of 100, where zero corresponds to the expected value.

The other types of complaint were less likely than the chronic illnesses to appear in the available records. From the nature of the survey questions, and from what we know of clinicians' priorities when they make record entries, it seems probable that the discrepancies can be attributed at least as much to the shortcomings of the medical records when judged as instruments for this kind of research as to the vagaries of patients' responses. The abstractors found it difficult to judge from the written record how well patients could carry out their normal activities and to what extent their eyeglasses and hearing aids were correcting the underlying impairments. Furthermore, they found that symptoms were not always recorded in detail unless the patient had taken the optional multiphasic examination.

The index of agreement for the whole set of dichotomous questions on physical complaints was 37, and for the fourfold ratings on the Physical Health Spectrum, 34.

Although the ratings of elderly patients on the Physical Health Spectrum were somewhat less reliable, they agreed better with the medical records than did those of younger patients. Part of the explanation is doubtless that the older people had relatively detailed records at Kaiser. The survey responses of the younger patients resemble those of patients whose records are incomplete—both groups tended to report more serious problems than their records showed.

Summary

The Human Population Laboratory of the California State Department of Public Health conducted two methodological studies to determine how consistently people answer questions about their health when a survey is repeated

after a short interval and how closely information collected by survey agrees with that obtained from clinical records.

In a 1965 baseline survey, data on chronic physical complaints were collected by self-administered questionnaires from a probability sample of adults in Alameda County. Validity, or extent of agreement between the survey information and that obtained from clinical records, was investigated by a record check on respondents in the baseline survey who had been receiving care through a prepaid health plan.

For 739 baseline survey respondents, 54 percent of the chronic conditions reported by questionnaire were reported in their clinical records. Other types of complaint were less likely to appear in the clinical records than in the questionnaires.

Another sample of 1,530 respondents was used in 1968 to measure reproducibility of interview information. These respondents completed two identical questionnaires, self-administered about a week apart, on their physical health—disabilities, chronic illnesses, impairments, and symptoms.

For the whole set of 33 dichotomous questions on physical complaints, 96 percent of the responses were consistent. Chronic conditions were reported more reliably than other types of complaint, and negative answers were reported more reliably than affirmative answers.

REFERENCES

- (1) U.S. National Health Survey: A study of special purpose medical-history techniques. PHS Publication No. 584, ser. D, No. 1. U.S. Government Printing Office, Washington, D.C., 1960.
- (2) Bailar, B. A.: Recent research in reinterview procedures. *J Amer Stat Assoc* 63: 41-63, March 1968.
- (3) Balamuth, E.: Health interview responses compared with medical records. PHS Publication No. 1000, ser. 2, No. 7. U.S. Government Printing Office, Washington, D.C., July 1965.
- (4) Madow, W. G.: Interview data on chronic conditions compared with information derived from medical records. PHS Publication No. 1000, ser. 2, No. 23. U.S. Government Printing Office, Washington, D.C., May 1967.
- (5) Trussel, R. E., and Elinson, J.: Chronic illness in a rural area. Harvard University Press, Cambridge, Mass., 1959.
- (6) Hochstim, J. R.: A critical comparison of three

- strategies of collecting data from households. *J Amer Stat Assoc* 62: 976-980, September 1967.
- (7) California State Department of Public Health: Alameda County population 1965. Human Population Laboratory, series A, No. 7, Berkeley, April 1966.
- (8) Hochstim, J. R., and Renne, K. S.: Reliability of response in a sociomedical population study. *Public Opinion Quart.* In press.
- (9) Columbia University School of Public Health and

Administrative Medicine: Family medical care under three types of health insurance. Foundation on Employee Health, Medical Care and Welfare, Inc., New York, 1962.

Tearsheet Requests

Dr. Joseph R. Hochstim, Human Population Laboratory, California State Department of Public Health, 2151 Berkeley Way, Berkeley, Calif. 94704

Education Notes

Ph.D. Program in Urban and Regional Planning. The University of Michigan is offering an interdisciplinary program leading to a doctoral degree in urban and regional planning.

The program is under the general direction of the program committee on urban and regional planning, with faculty representatives from eight schools and colleges of the university. The course of study emphasizes both analysis of urban and regional processes and methods for generating, selecting, and implementing desired changes in those processes. Before a doctoral dissertation proposal is approved, the student must demonstrate competence in three of the following: environmental design and resource appraisal, analytic tools for urban and regional analysis (required), social group interaction processes, economic development of urban and regional areas, and governmental planning process.

All students are expected to participate in a first-year seminar, and those who have completed qualifying examinations will enroll in the research forum for at least one term. In addition, all students obtaining the Ph.D. must complete two full-time terms of relevant work experience or their equivalent (working half-time for four terms would fulfill the requirement). This work may be some combination of research independent of the doctoral dissertation, field or practicum experience, or teaching at the college level. Such experience can be obtained on or off the University of Michigan campus.

Fellowships and grants are awarded on a competitive basis. Some include stipend and tuition, others stipend only or tuition reduced to in-State status for out-of-State residents. Various opportunities for employment on research projects exist and normally result in greater financial support than for fellowship awards alone. Loans and deferred payment plans are available from the university. Re-

quests for financial assistance must be received before February 1.

Admission is open to highly qualified and motivated college graduates. The admissions committee of the program requires submission of a statement of intent, which will also help guide the student's advisers as to appropriate course work if admitted; three letters of recommendation; and examples of previous research or practical experience, if applicable. While admission is normally for the fall terms, applicants may, in some cases, begin their studies at other times of the year.

Forms for admission and financial support may be obtained from the Admissions Office, Horace H. Rackham School of Graduate Studies, University of Michigan, Ann Arbor, Mich. 48104.

Fellowships in Environmental Health. The Harvard School of Public Health has announced the availability of fellowships for U.S. citizens interested in graduate training in environmental health.

Specialized training is offered in air pollution control, radiological health, industrial hygiene, toxicology, radiobiology, environmental and respiratory physiology, aerospace health and safety, and occupational medicine.

Fellowships cover tuition and fees plus a monthly stipend depending on the academic degree held by the student when he enrolls. At the post-bachelor degree level, the stipend is about \$250 per month. An additional \$500 per year is provided for each dependent.

Professional personnel in State and local agencies, industrial staff members with interests in public health, and students completing bachelor degree programs in either the biological or physical sciences are encouraged to apply. Applications should be submitted by April 1, 1971.

Further information may be obtained from Dr. Dade W. Moeller, Kresge Center for Environmental Health, Harvard School of Public Health, 665 Huntington Ave., Boston, Mass. 02115.